

# Inferência Estatística II

# **Análise Bivariada**

Prof. Dr. Juliano van Melis

Parte II

## Comparação de duas médias

- Usando o R
- Dimensionamento da amostra para testes de diferenças
- Teste t student para diferença de duas médias provenientes de amostras independentes
- Teste t student para diferença de duas amostras relacionadas

## Comparação de proporções

- teste Z para diferença entre duas proporções
- teste Qui-quadrado para Aderência
- teste Qui-quadrado para Independência
- teste Qui-quadrado para Homogeneidade

# Parte 2

# Conteúdo

## Testes não-paramétricos

- Teste dos Sinais
- Teste Mann-Witney
- Teste de Wilcoxon



Análise Bivariada

# COMPARAÇÃO DE DUAS MÉDIAS

Objetivo: Comparar **DUAS** populações/amostras

**Segue distribuição Normal?**

**SIM**

Parte 1

# Objetivo: Comparar **DUAS** populações/amostras

**Segue distribuição Normal?**

**SIM**

**NÃO**

**É possível transformar os dados?**

→ Transformação Box-Cox

$Y^\lambda$ , sendo  $\lambda \in \{-2, -1, 1, 2\}$  e  $\log(Y)$

→ Quando, Como, Porque e Porque não:

<http://support.sas.com/resources/papers/proceedings12/430-2012.pdf>

→ Não use log em dados de contagem:

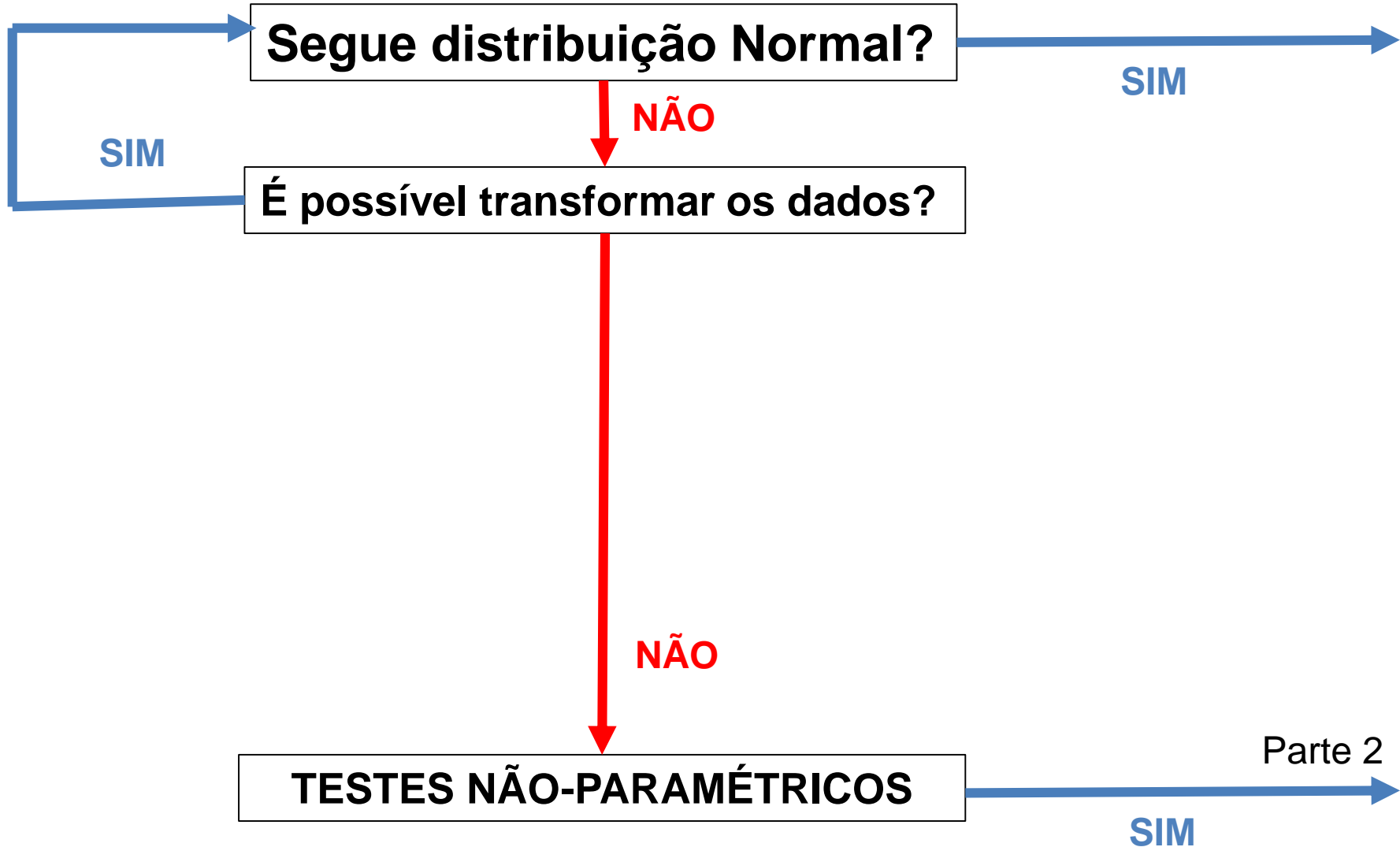
<http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2010.00021.x/pdf>

Objetivo: Comparar **DUAS** populações/amostras





Objetivo: Comparar **DUAS** populações/amostras



# Objetivo: Comparar **DUAS** populações/amostras

Parte 1

**Segue distribuição Normal?**



**SIM**

**Quer comparar as médias?**

# Objetivo: Comparar **DUAS** populações/amostras

Parte 1

**Segue distribuição Normal?**



**SIM**

**Quer comparar as médias?**

**NÃO**

Outros testes



# Objetivo: Comparar **DUAS** populações/amostras

**Segue distribuição Normal?**



**SIM**

**Quer comparar as médias?**

**NÃO**

Outros testes

**SIM**

**Amostras são dependentes?**

# Objetivo: Comparar **DUAS** populações/amostras

**Segue distribuição Normal?**

**SIM**

**Quer comparar as médias?**

**NÃO**

Outros testes

**SIM**

**Amostras são dependentes?**

**SIM**

**Teste de t de Student  
PAREADO  
g.l. =  $N - 1$**

# Objetivo: Comparar **DUAS** populações/amostras

Segue distribuição Normal?

SIM

Quer comparar as médias?

NÃO

Outros testes

Amostras são dependentes?

SIM

SIM

Teste de t de Student  
**PAREADO**  
g.l. = N - 1

$$T = \frac{\bar{D}}{\sqrt{\frac{S_D^2}{n}}}$$

$$D = \text{Medida}_{\text{depois}} - \text{Medida}_{\text{antes}}$$

$$S_D^2 = \frac{1}{n - 1} \sum_{i=1}^n (D_i - \bar{D})^2.$$

# Objetivo: Comparar **DUAS** populações/amostras

**Segue distribuição Normal?**

**SIM**

**Quer comparar as médias?**

**NÃO**

Outros testes

**SIM**

**Amostras são dependentes?**

**SIM**

**Teste de t de Student  
PAREADO  
g.l. =  $N - 1$**

**NÃO**

**As variâncias são iguais?**

**SIM**

**Teste de t de Student  
g.l. =  $N_1 + N_2 - 2$**

**Bartlett ou Levene  
 $p > 0.05$**

# Objetivo: Comparar **DUAS** populações/amostras

Segue distribuição Normal?

SIM

Quer comparar as médias?

NÃO

Outros testes

Amostras são dependentes?

SIM

Teste de t de Student  
**PAREADO**  
g.l. =  $N - 1$

NÃO

As variâncias são iguais?

SIM

Teste de t de Student  
g.l. =  $N_1 + N_2 - 2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}$  : média da amostra x  
 $s_x$  : desvio padrão amostra x  
 $n_1$  : número de amostras x



# Objetivo: Comparar **DUAS** populações/amostras

Segue distribuição Normal?

SIM

Quer comparar as médias?

NÃO

Outros testes

Amostras são dependentes?

SIM

Teste de t de Student  
**PAREADO**  
g.l. =  $N - 1$

NÃO

As variâncias são iguais?

SIM

Teste de t de Student  
g.l. =  $N_1 + N_2 - 2$

NÃO

Teste de t de Student  
g.l. =  $v = d.f.$

$$d.f. = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

# Objetivo: Comparar **DUAS** populações/amostras

Segue distribuição Normal?

SIM

Quer comparar as médias?

NÃO

Outros testes

SIM

Amostras são dependentes?

SIM

Teste de t de Student  
**PAREADO**  
g.l. =  $N - 1$

NÃO

As variâncias são iguais?

SIM

Teste de t de Student  
g.l. =  $N_1 + N_2 - 2$

NÃO

Teste de t de Student  
g.l. =  $v$

$$\nu \approx \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^2 \nu_1} + \frac{s_2^4}{N_2^2 \nu_2}} \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

# PRESSUPOSTOS

## **Pressupostos do Teste**

1. Distribuição normal dos dados
2. Variâncias iguais (Homocedasticidade)
  - Ajuste no grau de liberdade
3. As observações devem ser independentes
  - Teste com dependência

# PROVIDÊNCIAS

**(i) Estabelecer as hipóteses (nula e alternativa);**

Se bicaudal:

**H0:** As médias X1 e X2 são **iguais**;

**H1:** As médias X1 e X2 são **diferentes**;

Se unicaudal:

**H0:** A média X1 é **menor/menor** que a média de X2;

**H1:** A média X1 é **maior/menor** que a média de X2;

**(ii) Verificar pressupostos;**

**(iii) Estabelecer  $\alpha$ ;**

Essa é a probabilidade de erro tipo I aceitável , também chamada de nível de significância). **Normalmente o valor é  $\alpha=0.05$ .** → Associado ao  $T_{crítico}$

**(iv) Calcular a estatística do teste (valor do  $T_{calculado}$ );**

**(v) Determine o p-valor e comparar com o  $T_{crítico}$ .**

Se **p-valor < 0.05**: Rejeite H0 (aceitamos H1)

$$|t_{crítico}| < |t_{calculado}|$$

Se **p-valor > 0.05**: Não rejeite H0 (aceitamos a H0)

$$|t_{crítico}| > |t_{calculado}|$$

**(vi) Concluir e reportar por extenso.**

# EXEMPLO

Deseja-se verificar se existe diferença entre os salários pagos a engenheiros que atuam na região Sul e Sudeste do país através de um teste de hipóteses. Para isso, selecionaram-se aleatoriamente 30 engenheiros da região Sul e, com base em seus salários anuais, determinou-se a média como sendo R\$ 46.720,00 com desvio padrão de R\$ 14.700,00. O mesmo procedimento foi adotado para 35 engenheiros da região Sudeste, obtendo-se média de R\$ 51.910,00 e desvio padrão de R\$ 16.200,00. Assuma um  $\alpha = 5\%$  e considere que a normalidade e a homogeneidade foram respeitadas.

# EXEMPLO

- (i) Estabelecer Hipóteses:**
- (ii) Verificar pressupostos:**
- (iii) Estabelecer  $\alpha$**
- (iv) Calcular a estatística do teste**
- (v) Determine p-valor, comparando com  $\alpha$ , comparar  $t_{\text{calculado}}$  e  $t_{\text{crítico}}$**
- (vi) Concluir**

# EXEMPLO

Deseja-se verificar **se existe diferença entre os salários pagos a engenheiros que atuam na região Sul e Sudeste do país** através de um teste de hipóteses. Para isso, selecionaram-se aleatoriamente 30 engenheiros da região Sul e, com base em seus salários anuais, determinou-se a média como sendo R\$ 46.720,00 com desvio padrão de R\$ 14.700,00. O mesmo procedimento foi adotado para 35 engenheiros da região Sudeste, obtendo-se média de R\$ 51.910,00 e desvio padrão de R\$ 16.200,00. Assuma um  $\alpha = 5\%$  e considere que a normalidade e a homogeneidade foram respeitadas.

# EXEMPLO

**(i) Estabelecer Hipóteses:**

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

**(ii) Verificar pressupostos:**

**(iii) Estabelecer  $\alpha$**

**(iv) Calcular a estatística do teste**

**(v) Determine p-valor, comparando com  $\alpha$ , comparar  $t_{\text{calculado}}$  e  $t_{\text{crítico}}$**

**(vi) Concluir**



# EXEMPLO

**(i) Estabelecer Hipóteses:**

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

**(ii) Verificar pressupostos:**

- Segue Normal
- Variâncias iguais (homocedásticos)
- Observações independentes

**(iii) Estabelecer  $\alpha$**

**(iv) Calcular a estatística do teste**

**(v) Determine p-valor, comparando com  $\alpha$ , comparar  $t_{\text{calculado}}$  e  $t_{\text{crítico}}$**

**(vi) Concluir**

# EXEMPLO

Deseja-se verificar se existe diferença entre os salários pagos a engenheiros que atuam na região Sul e Sudeste do país através de um teste de hipóteses. Para isso, selecionaram-se aleatoriamente 30 engenheiros da região Sul e, com base em seus salários anuais, determinou-se a média como sendo R\$ 46.720,00 com desvio padrão de R\$ 14.700,00. O mesmo procedimento foi adotado para 35 engenheiros da região Sudeste, obtendo-se média de R\$ 51.910,00 e desvio padrão de R\$ 16.200,00. **Assuma um  $\alpha = 5\%$  e considere que a normalidade e a homogeneidade foram respeitadas.**

# EXEMPLO

## **(i) Estabelecer Hipóteses:**

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

## **(ii) Verificar pressupostos:**

- Segue Normal
- Variâncias iguais (homocedásticos)
- Observações independentes

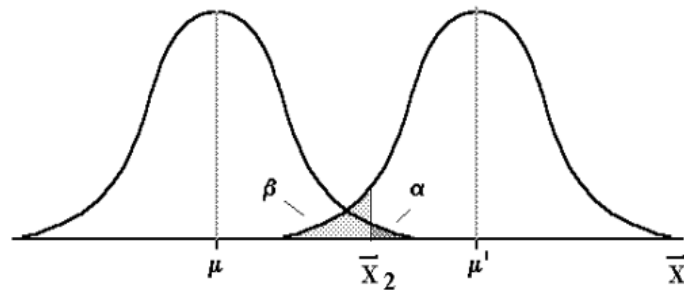
## **(iii) Estabelecer $\alpha$**

- 5% (bicaudal)  $\rightarrow t_{\text{crítico}} =$

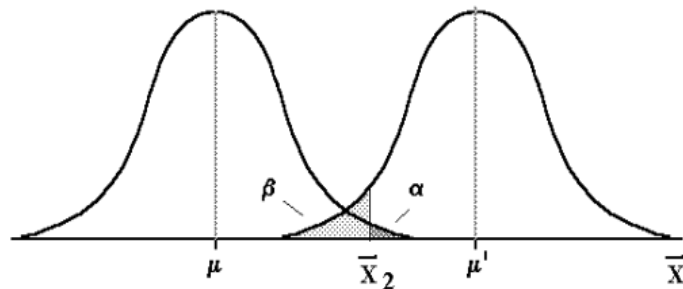
## **(iv) Calcular a estatística do teste**

## **(v) Determine p-valor, comparando com $\alpha$ , comparar $t_{\text{calculado}}$ e $t_{\text{crítico}}$**

## **(vi) Concluir**



$\alpha$	0,200	0,150	0,100	0,050	0,025	0,020	0,015	0,010	0,005
g.l.1	1,37638	1,96261	3,07768	6,31375	12,70615	15,89447	21,20505	31,82096	63,65590
2	1,06066	1,38621	1,88562	2,91999	4,30266	4,84873	5,64280	6,96455	9,92499
3	0,97847	1,24978	1,63775	2,35336	3,18245	3,48191	3,89606	4,54071	5,84085
4	0,94096	1,18957	1,53321	2,13185	2,77645	2,99853	3,29763	3,74694	4,60408
5	0,91954	1,15577	1,47588	2,01505	2,57058	2,75651	3,00288	3,36493	4,03212
6	0,90570	1,13416	1,43976	1,94318	2,44691	2,61224	2,82893	3,14267	3,70743
7	0,89603	1,11916	1,41492	1,89458	2,36462	2,51675	2,71457	2,99795	3,49948
8	0,88889	1,10815	1,39682	1,85955	2,30601	2,44899	2,63381	2,89647	3,35538
9	0,88340	1,09972	1,38303	1,83311	2,26216	2,39844	2,57381	2,82143	3,24984
10	0,87906	1,09306	1,37218	1,81246	2,22814	2,35931	2,52749	2,76377	3,16926
11	0,87553	1,08767	1,36343	1,79588	2,20099	2,32814	2,49067	2,71808	3,10582
12	0,87261	1,08321	1,35622	1,78229	2,17881	2,30272	2,46070	2,68099	3,05454
13	0,87015	1,07947	1,35017	1,77093	2,16037	2,28160	2,43585	2,65030	3,01228
14	0,86805	1,07628	1,34503	1,76131	2,14479	2,26378	2,41490	2,62449	2,97685
15	0,86624	1,07353	1,34061	1,75305	2,13145	2,24854	2,39701	2,60248	2,94673
16	0,86467	1,07114	1,33676	1,74588	2,11990	2,23536	2,38155	2,58349	2,92079
17	0,86328	1,06903	1,33338	1,73961	2,10982	2,22384	2,36805	2,56694	2,89823
18	0,86205	1,06717	1,33039	1,73406	2,10092	2,21370	2,35618	2,55238	2,87844
19	0,86095	1,06551	1,32773	1,72913	2,09302	2,20470	2,34565	2,53948	2,86094
20	0,85996	1,06402	1,32534	1,72472	2,08596	2,19666	2,33625	2,52798	2,84534
21	0,85907	1,06267	1,32319	1,72074	2,07961	2,18943	2,32779	2,51765	2,83137
22	0,85827	1,06145	1,32124	1,71714	2,07388	2,18289	2,32016	2,50832	2,81876
23	0,85753	1,06034	1,31946	1,71387	2,06865	2,17696	2,31323	2,49987	2,80734
24	0,85686	1,05932	1,31784	1,71088	2,06390	2,17155	2,30692	2,49216	2,79695
25	0,85624	1,05838	1,31635	1,70814	2,05954	2,16659	2,30113	2,48510	2,78744
26	0,85567	1,05752	1,31497	1,70562	2,05553	2,16203	2,29581	2,47863	2,77872



$\alpha$	0,200	0,150	0,100	0,050	0,025	0,020	0,015	0,010	0,005
27	0,85514	1,05673	1,31370	1,70329	2,05183	2,15782	2,29092	2,47266	2,77068
28	0,85465	1,05599	1,31253	1,70113	2,04841	2,15394	2,28638	2,46714	2,76326
29	0,85419	1,05530	1,31143	1,69913	2,04523	2,15033	2,28218	2,46202	2,75639
30	0,85377	1,05466	1,31042	1,69726	2,04227	2,14697	2,27827	2,45726	2,74998
35	0,85201	1,05202	1,30621	1,68957	2,03011	2,13316	2,26219	2,43772	2,72381
40	0,85070	1,05005	1,30308	1,68385	2,02107	2,12291	2,25027	2,42326	2,70446
45	0,84968	1,04852	1,30065	1,67943	2,01410	2,11500	2,24109	2,41212	2,68959
50	0,84887	1,04729	1,29871	1,67591	2,00856	2,10872	2,23378	2,40327	2,67779
60	0,84765	1,04547	1,29582	1,67065	2,00030	2,09936	2,22292	2,39012	2,66027
70	0,84679	1,04417	1,29376	1,66692	1,99444	2,09273	2,21523	2,38080	2,64790
80	0,84614	1,04319	1,29222	1,66413	1,99007	2,08778	2,20949	2,37387	2,63870
90	0,84563	1,04244	1,29103	1,66196	1,98667	2,08394	2,20504	2,36850	2,63157
100	0,84523	1,04184	1,29008	1,66023	1,98397	2,08088	2,20150	2,36421	2,62589
110	0,84490	1,04134	1,28930	1,65882	1,98177	2,07839	2,19860	2,36072	2,62127
120	0,84463	1,04093	1,28865	1,65765	1,97993	2,07631	2,19620	2,35783	2,61742
140	0,84420	1,04029	1,28763	1,65581	1,97706	2,07306	2,19244	2,35328	2,61140
160	0,84387	1,03980	1,28686	1,65443	1,97490	2,07063	2,18962	2,34988	2,60690
180	0,84362	1,03943	1,28627	1,65336	1,97323	2,06874	2,18743	2,34724	2,60341
200	0,84342	1,03913	1,28580	1,65251	1,97189	2,06723	2,18569	2,34513	2,60063
$\infty$	0,84198	1,03697	1,28240	1,64638	1,96234	2,05643	2,17319	2,33008	2,58075

# EXEMPLO

## (i) Estabelecer Hipóteses:

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

## (ii) Verificar pressupostos:

- Segue Normal
- Variâncias iguais (homocedásticos)
- Observações independentes

## (iii) Estabelecer $\alpha$

- 5% (bicaudal)  $\rightarrow t_{\text{crítico}} = \pm 2,0003$

## (iv) Calcular a estatística do teste

$X_1 =$

$X_2 =$

$s_1 =$

$s_2 =$

$n_1 =$

$n_2 =$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

## (v) Determine p-valor, comparando com $\alpha$ , comparar $t_{\text{calculado}}$ e $t_{\text{crítico}}$

## (vi) Concluir

# EXEMPLO

Deseja-se verificar se existe diferença entre os salários pagos a engenheiros que atuam na região Sul e Sudeste do país através de um teste de hipóteses. Para isso, selecionaram-se aleatoriamente 30 engenheiros da região Sul e, com base em seus salários anuais, determinou-se a média como sendo R\$ 46.720,00 com desvio padrão de R\$ 14.700,00. O mesmo procedimento foi adotado para 35 engenheiros da região Sudeste, obtendo-se média de R\$ 51.910,00 e desvio padrão de R\$ 16.200,00. Assuma um  $\alpha = 5\%$  e considere que a normalidade e a homogeneidade foram respeitadas.

# EXEMPLO

## (i) Estabelecer Hipóteses:

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

## (ii) Verificar pressupostos:

- Segue Normal
- Variâncias iguais (homocedásticos)
- Observações independentes

## (iii) Estabelecer $\alpha$

- 5% (bicaudal)  $\rightarrow t_{\text{crítico}} = \pm 2,0003$

## (iv) Calcular a estatística do teste

$$X_1 = 51.910,00$$

$$X_2 = 46.720,00$$

$$s_1 = 16.200,00$$

$$s_2 = 14.700,00$$

$$n_1 = 35$$

$$n_2 = 30$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

## (v) Determine p-valor, comparando com $\alpha$ , comparar $t_{\text{calculado}}$ e $t_{\text{crítico}}$

$$t_{\text{crítico}} =$$

$$t_{\text{calculado}} =$$

## (vi) Concluir



# EXEMPLO

## (i) Estabelecer Hipóteses:

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

## (ii) Verificar pressupostos:

- Segue Normal
- Variâncias iguais (homocedásticos)
- Observações independentes

## (iii) Estabelecer $\alpha$

- 5% (bicaudal)  $\rightarrow t_{\text{crítico}} = \pm 2,0003$

## (iv) Calcular a estatística do teste

$$X_1 = 51.910,00$$

$$X_2 = 46.720,00$$

$$s_1 = 16.200,00$$

$$s_2 = 14.700,00$$

$$n_1 = 35$$

$$n_2 = 30$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

## (v) Determine p-valor, comparando com $\alpha$ , comparar $t_{\text{calculado}}$ e $t_{\text{crítico}}$

$$t_{\text{crítico}} = \pm 2,0$$

$$t_{\text{calculado}} = 1,354$$

## (vi) Concluir

# EXEMPLO

## (i) Estabelecer Hipóteses:

Hipótese Nula: Média População 1 = Média População 2

Hipótese Nula: Média População 1  $\neq$  Média População 2

## (ii) Verificar pressupostos:

- Segue Normal
- Variâncias iguais (homocedásticos)
- Observações independentes

## (iii) Estabelecer $\alpha$

- 5% (bicaudal)  $\rightarrow t_{\text{crítico}} = \pm 2,0003$

## (iv) Calcular a estatística do teste

$$X_1 = 51.910,00$$

$$X_2 = 46.720,00$$

$$s_1 = 16.200,00$$

$$s_2 = 14.700,00$$

$$n_1 = 35$$

$$n_2 = 30$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$



Rejeita H0



Não rejeita H0

## (v) Determine p-valor, comparando com $\alpha$ , comparar $t_{\text{calculado}}$ e $t_{\text{crítico}}$

$$t_{\text{crítico}} = \pm 2,0$$

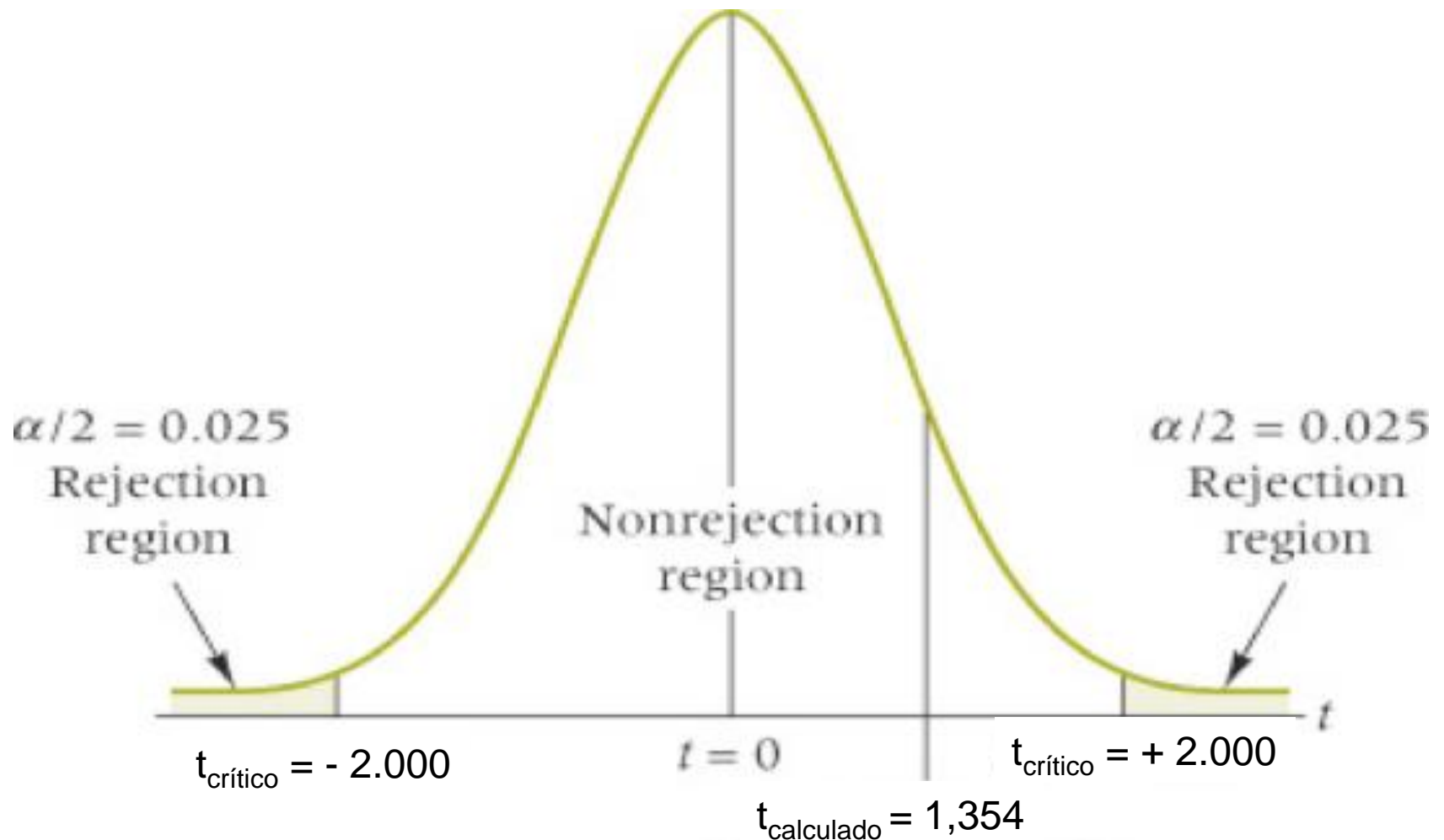
$$t_{\text{calculado}} = 1,354$$

## (vi) Concluir

# EXEMPLO

$$t_{\text{crítico}} = \pm 2,0$$

$$t_{\text{calculado}} = 1,354$$



# EXEMPLO

(vi) Conclusão redigida

Para um nível de significância de 5%, **não rejeitamos a  $H_0$** , ou seja, os dados proporcionaram evidência para concluir que a média salarial de engenheiros na região Sudeste (média = R\$ 51.910,00, erro padrão = R\$ 2.738,30) é igual à média salarial de engenheiros na região Sul (média = 46.720,00, erro padrão = 2.683,84) ( $t_{(63)} = 1,354$ ;  $p > 0,10$ ).

# EXERCÍCIOS

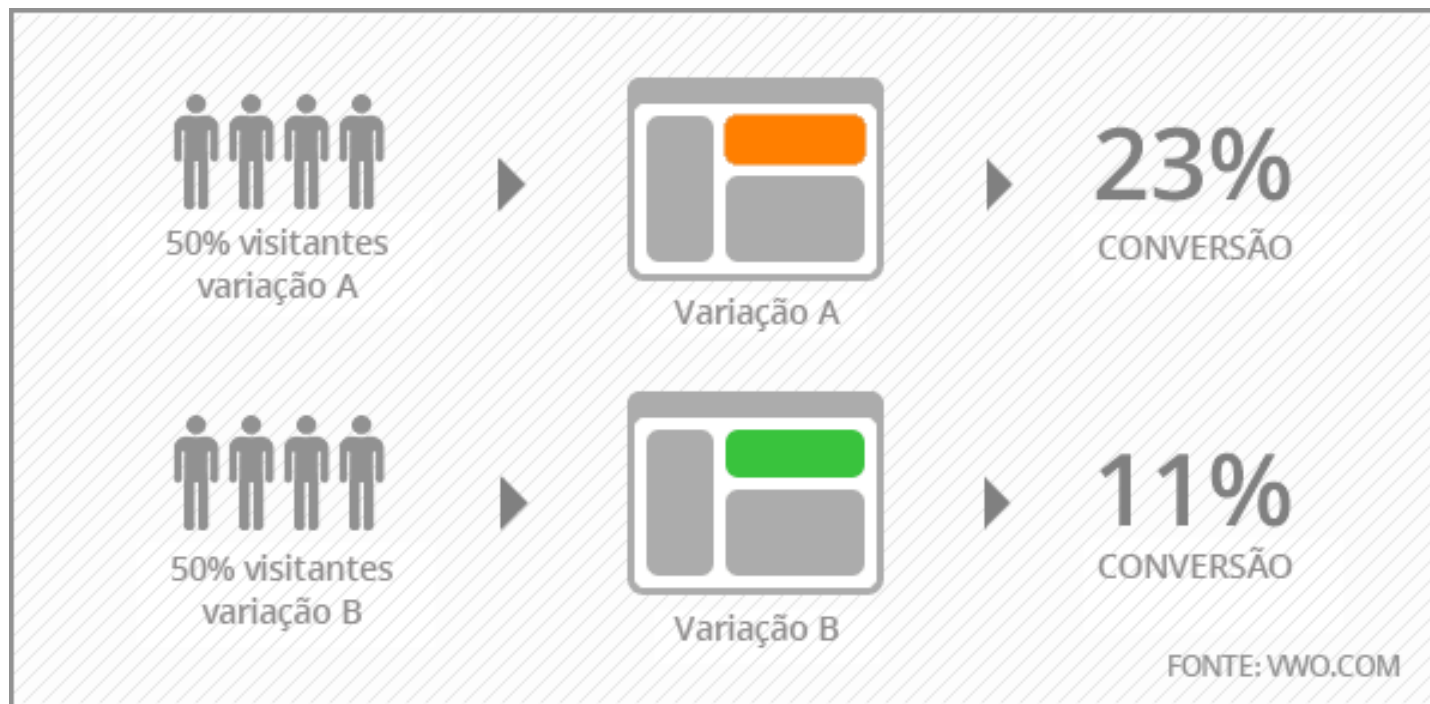
→ Fazer sequência do Exercício 3

# EXERCÍCIOS

→ Fazer no R



Abrir arquivo “teste\_t.Rmd”



Análise Bivariada

# COMPARAÇÃO DE PROPORÇÕES

vs



**Grupo A**  
480/500



**Grupo B**  
400/500

1.  $H_0 : p_A = p_B$

2.  $H_0 : p_A \leq p_B$

3.  $H_0 : p_A \geq p_B$

1.  $H_a : p_A \neq p_B$

2.  $H_a : p_A > p_B$

3.  $H_a : p_A < p_B$

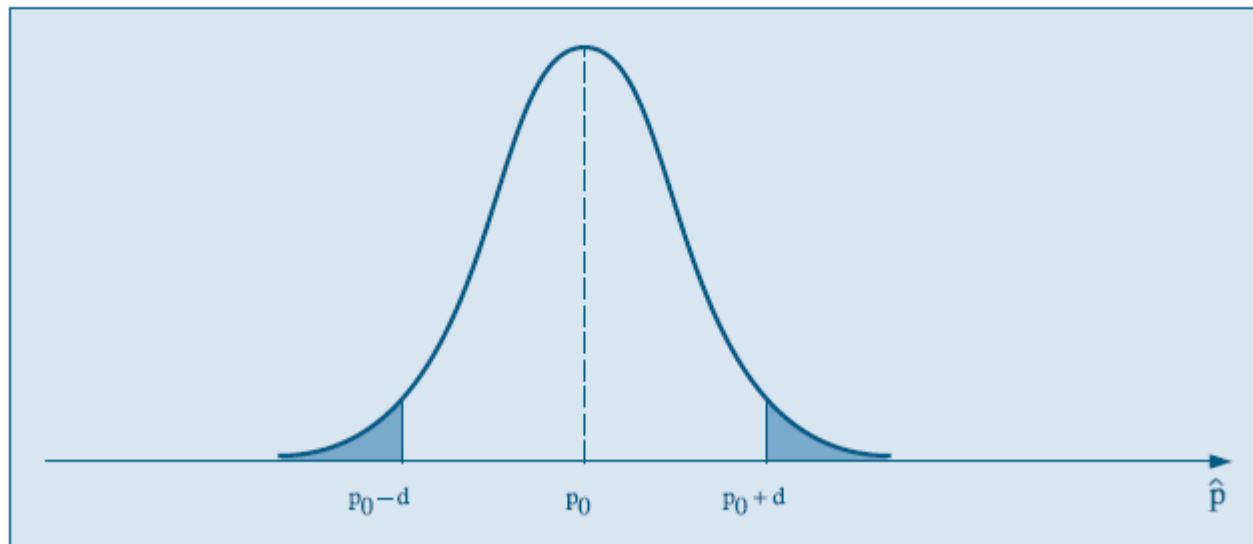
bicaudal

unicaudal



$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

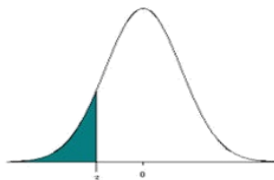
**Figura 12.6:** Região crítica para o teste  $H_0 : p = p_0$  vs  $H_1 : p \neq p_0$ .



$$z = \frac{p_A - p_B}{\sqrt{pq/n_A + pq/n_B}}$$

Table of Standard Normal Probabilities for Negative Z-scores

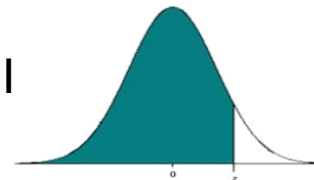
Tabela I



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table of Standard Normal Probabilities for Positive Z-scores

Tabela II



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Note that the probabilities given in this table represent the area to the LEFT of the z-score.

The area to the RIGHT of a z-score = 1 – the area to the LEFT of the z-score

```
prop.test(x, n,  
          p = NULL,  
          alternative = "two.sided",  
          correct = TRUE)
```



**x:** Vetor - Contagem do número de sucessos

**n:** Vetor – Contagem do número de tentativas

**alternative:** Caracter – especificando a hipótese alternativa

**correct:** Lógico – Se correção de Yates deve ser realizada

```
> prop.test(x = c(490,400),  
+           n = c(500, 500))
```

2-sample test for equality of proportions with continuity correction

```
data:  c(490, 400) out of c(500, 500)  
X-squared = 80.909, df = 1, p-value < 2.2e-16  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 0.1408536 0.2191464  
sample estimates:  
prop 1 prop 2  
 0.98   0.80
```

```
> |
```



## → Tamanho amostral

$$n = \frac{Z_{gc}^2 \cdot p \cdot q}{e^2}$$

→  $Z_{gc}$ : Valor de z relativo ao grau de confiança (95% = 1.96)

→ e: erro absoluto (“para mais ou para meno”)

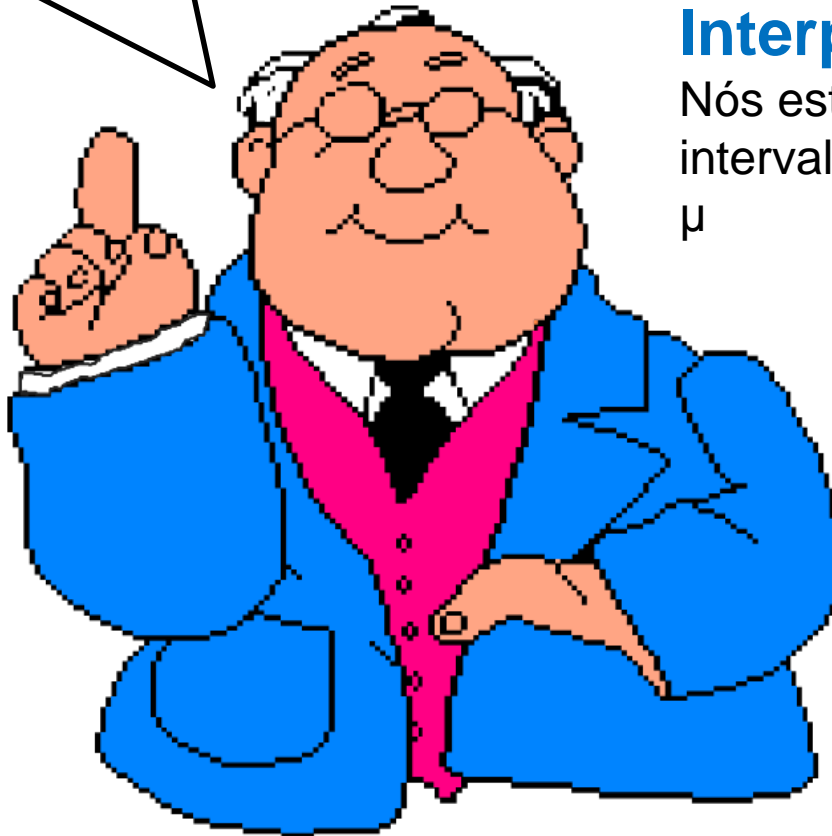
→ p: probabilidade do evento

→ q: probabilidade do não evento (1-p)



## → Tamanho amostral

Tenho 95% de confiança  
que está entre 10% e 30%



### Interpretação prática:

Nós estamos  $100(1-\alpha)\%$  confiantes que o intervalo de confiança contém o valor de  $\mu$

## EXERCÍCIO

1. Os produtores de um programa de televisão pretendem modificá-lo se for assistido regularmente por menos de um quarto dos possuidores de televisão. Uma pesquisa encomendada a uma empresa especializada mostrou que, de 400 famílias entrevistadas, 80 assistem ao programa regularmente. Com base nos dados, qual deve ser a decisão dos produtores?

Temos uma população  $P$  e queremos verificar se ela segue uma distribuição especificada  $P_0$ , isto é, queremos testar a hipótese  $H_0 : P = P_0$ .

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$



Temos uma população  $P$  e queremos verificar se ela segue uma distribuição especificada  $P_0$ , isto é, queremos testar a hipótese  $H_0 : P = P_0$ .

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$



**Tabela 14.1:** Resultados do lançamento de um dado 300 vezes.

Ocorrência ( $i$ )	1	2	3	4	5	6	Total
Freq. Observada ( $n_i$ )	43	49	56	45	66	41	300

Temos uma população  $P$  e queremos verificar se ela segue uma distribuição especificada  $P_0$ , isto é, queremos testar a hipótese  $H_0 : P = P_0$ .

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$



**Tabela 14.1:** Resultados do lançamento de um dado 300 vezes.

Ocorrência ( $i$ )	1	2	3	4	5	6	Total
Freq. Observada ( $n_i$ )	43	49	56	45	66	41	300
Freq. Esperada ( $n_i^*$ )							300

Temos uma população  $P$  e queremos verificar se ela segue uma distribuição especificada  $P_0$ , isto é, queremos testar a hipótese  $H_0 : P = P_0$ .

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$



**Tabela 14.1:** Resultados do lançamento de um dado 300 vezes.

Ocorrência ( $i$ )	1	2	3	4	5	6	Total
Freq. Observada ( $n_i$ )	43	49	56	45	66	41	300
Freq. Esperada ( $n_i^*$ )							300

$$1/6 \rightarrow n_{ij}^*/300$$

Temos uma população  $P$  e queremos verificar se ela segue uma distribuição especificada  $P_0$ , isto é, queremos testar a hipótese  $H_0 : P = P_0$ .

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$



**Tabela 14.1:** Resultados do lançamento de um dado 300 vezes.

Ocorrência ( $i$ )	1	2	3	4	5	6	Total
Freq. Observada ( $n_i$ )	43	49	56	45	66	41	300
Freq. Esperada ( $n_i^*$ )	50	50	50	50	50	50	300

Calcule o  $\chi^2$

# Hardy-Weinberg

Genetic structure of parent population

phenotypes



genotypes

$AA$

$Aa$

$aa$

number of moths  
(total = 500)

320

160

20

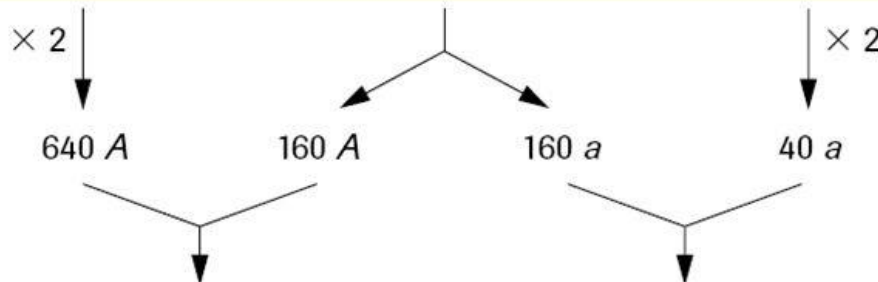
genotype frequencies

$$\frac{320}{500} = 0.64 \text{ } AA$$

$$\frac{160}{500} = 0.32 \text{ } Aa$$

$$\frac{20}{500} = 0.04 \text{ } aa$$

number of alleles  
in gene pool  
(total = 1000)



allele frequencies

$$\frac{800}{1000} = 0.8 \text{ } A$$

$$\frac{200}{1000} = 0.2 \text{ } a$$

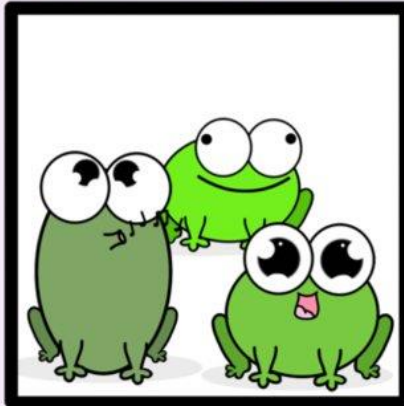
$p$  = frequency of  $A$  = 0.8

$q$  = frequency of  $a$  = 0.2

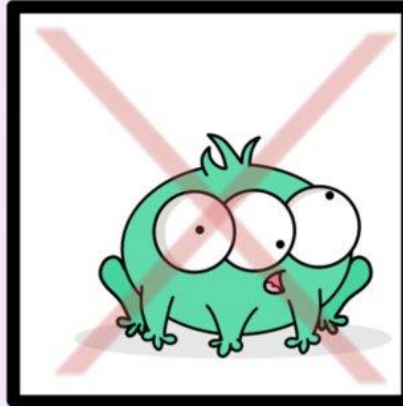
# Hardy-Weinberg

## Assumptions of Hardy-Weinberg Equilibrium

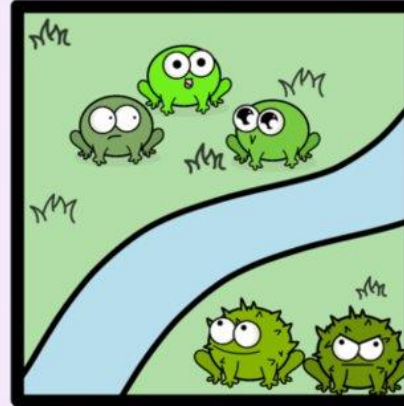
1. No selection



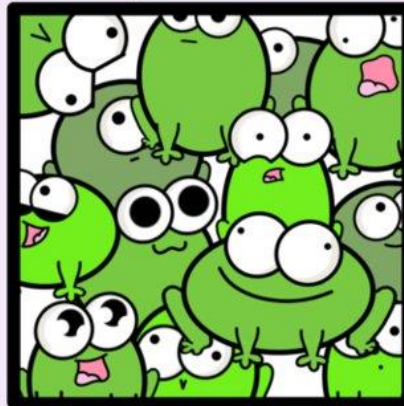
2. NO Mutation



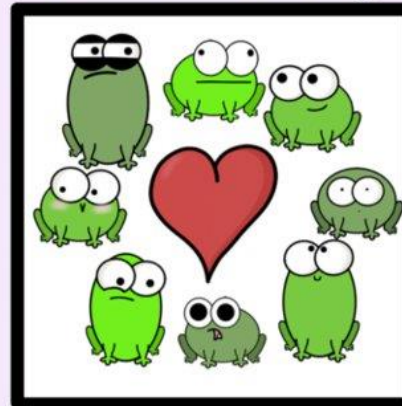
3. NO Migration



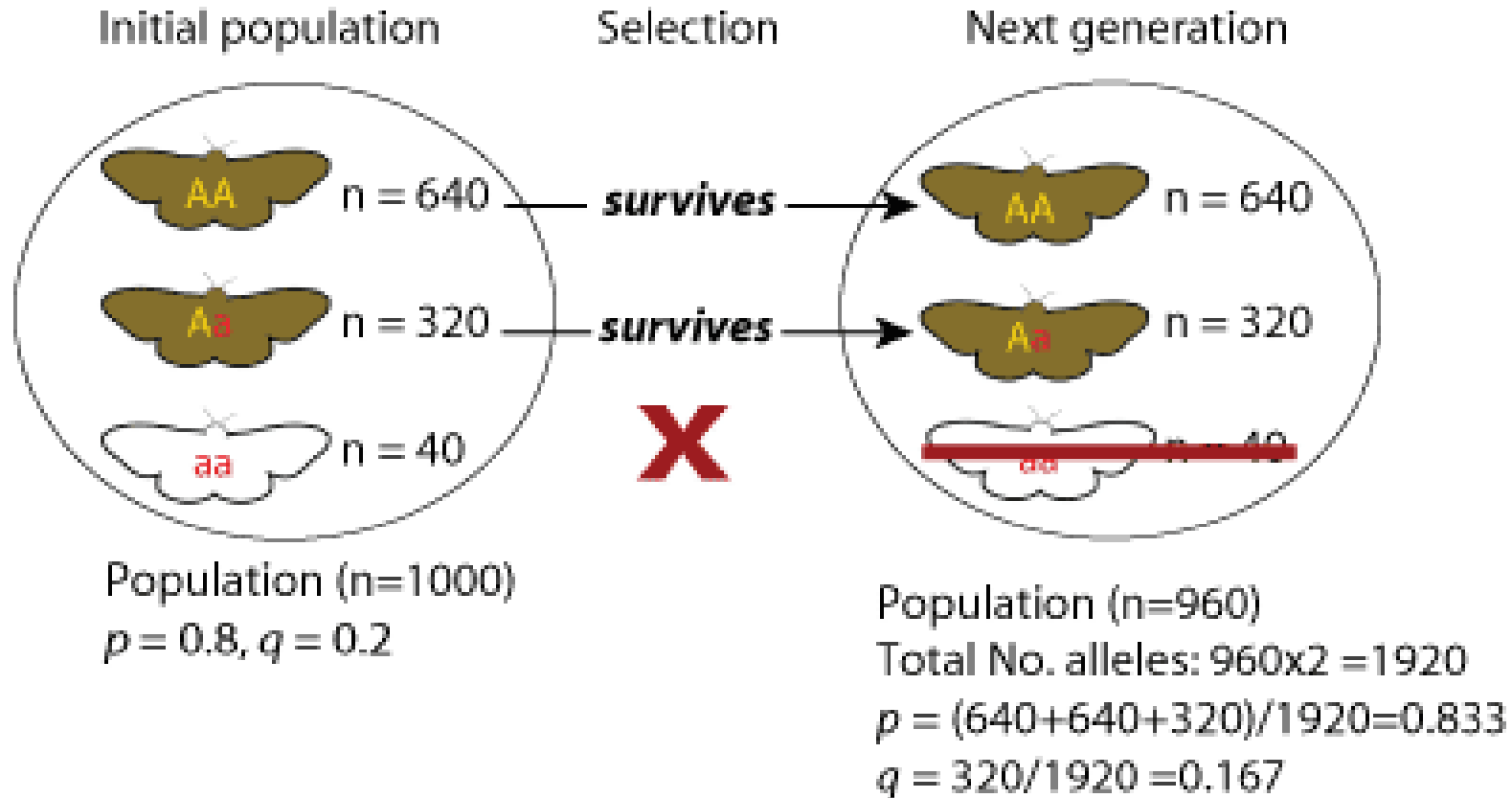
4. Large Population



5. Random Mating



# Hardy-Weinberg



# Hardy-Weinberg

	AA	Aa	aa	Total
Observado	$F_{AA}$	$F_{Aa}$	$F_{aa}$	$N = F_{AA} + F_{Aa} + F_{aa}$
Esperado	$p^2 N$	$2pqN$	$q^2 N$	$N$
Contribuição para $\chi^2$	$\frac{(F_{AA} - p^2 N)^2}{p^2 N}$	$\frac{(F_{Aa} - 2pqN)^2}{2pqN}$	$\frac{(F_{aa} - q^2 N)^2}{q^2 N}$	$\chi^2$

3 frequências  $\rightarrow$  g.l. = 3-1



# EXERCÍCIOS

Determine se a população a seguir encontra-se em equilíbrio de Hardy-Weinberg

98 AA;

50 Aa;

20 aa

N = 168

*Considere Qui-Quadrado crítico  
(associado à probabilidade de 5%) = 5,99*

## Probabilidade Condicional e Independência

Dois eventos, A e B de um mesmo espaço amostral são independentes quando a probabilidade de que eles ocorram simultaneamente, for igual ao produto de suas probabilidades individuais

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

$$P(A|B) = P(A)P(B)$$

A partir do resultado de um deles não é possível inferir nenhuma conclusão sobre o outro.

## “Teste” de Independência

Se  $P(A \cap B) \neq P(A) \cdot P(B)$ ,

podemos considerar que existe dependência (ou associação) entre eventos.

Por exemplo:

$$P(H - \text{hipertensos}) = 23\% = 0.23$$

$$P(M - \text{hipertensas}) = 18\% = 0.18$$

$$P(\text{casais} - \text{hipertensos}) = 7.2\%$$

→ **Podemos considerar associação ou dependência?**

# EXERCÍCIOS

Realizou-se uma pesquisa com uma amostra de 400 frequentadores de um clube esportivo, sendo 150 mulheres e 250 homens, a fim de classificá-los de acordo com a modalidade esportiva preferida: vôlei, basquete ou tênis. Os dados coletados na pesquisa estão descritos na tabela a seguir:

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	75	25	50	150
HOMENS	40	150	60	250
TOTAL	115	175	110	400

a) Se uma pessoa é amostrada aleatoriamente, qual é a probabilidade desta pessoa ser uma mulher e praticar Vôlei?

# EXERCÍCIOS

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	75	25	50	150
HOMENS	40	150	60	250
TOTAL	115	175	110	400

Construa uma tabela com as frequências esperadas entre modalidades e gênero

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES				150
HOMENS				250
TOTAL	115	175	110	400

# EXERCÍCIOS

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	75	25	50	150
HOMENS	40	150	60	250
TOTAL	115	175	110	400

Construa uma tabela com as frequências esperadas entre modalidades e gênero

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	Pmulher*Volei	Pmulher*basq	Pmulher*tenis	150
HOMENS	Phomem*Volei	Phomem*basq	Phomem*tenis	250
TOTAL	115	175	110	400

# EXERCÍCIOS

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	75	25	50	150
HOMENS	40	150	60	250
TOTAL	115	175	110	400

Construa uma tabela com as frequências esperadas entre modalidades e gênero

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	$(150/400)*115$	$(150/400)*175$	$(150/400)*110$	150
HOMENS	$(250/400)*115$	$(250/400)*175$	$(250/400)*110$	250
TOTAL	115	175	110	400

# EXERCÍCIOS

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	75	25	50	150
HOMENS	40	150	60	250
TOTAL	115	175	110	400

Construa uma tabela com as frequências esperadas entre modalidades e gênero

GÊNERO	VÔLEI	BASQUETE	TÊNIS	TOTAL
MULHERES	43.13	65.625	41.25	150
HOMENS	71.88	109.375	68.75	250
TOTAL	115	175	110	400



# EXERCÍCIOS

**Calcule o valor do Qui-quadrado.**

**Considere o g.l. =  $(linhas - 1) * (colunas - 1)$ .**

$$\chi^2 = \sum \frac{(Obs - Esp)^2}{Esp}$$

# EXERCÍCIOS

Uma pesquisa sobre a qualidade de certo produto foi realizada enviando-se questionários a donas-de-casa pelo correio. Aventando-se a possibilidade de que os respondentes voluntários tenham um particular viés de respostas, fizeram-se mais duas tentativas com os não-respondentes. Os resultados estão indicados abaixo. Você acha que existe relação entre a resposta e o número de tentativas?

Opinião sobre o produto	Nº de donas-de-casa		
	1ª tentativa	2ª tentativa	3ª tentativa
Excelente	62	36	12
Satisfatório	84	42	14
Insatisfatório	24	22	24

## Teste de Homogeneidade

Verificar se uma variável aleatória se comporta de modo similar, ou homogêneo, em várias subpopulações. Em outras palavras, em um teste de Chi Quadrado de homogeneidade podemos testar a afirmação de que diferentes populações têm a mesma proporção de indivíduos com alguma característica.

## Teste de Independência

O teste de Chi Quadrado de independência é semelhante ao teste de Chi Quadrado de aderência, mas considera uma “lei oriunda da própria tabela de dados experimentais” a fim de avaliar se há ou não dependência entre duas variáveis. Quanto maior a dependência entre as duas variáveis, maior será o valor de  $\chi^2$ . Quando as duas variáveis são independentes, o valor de  $\chi^2$  tende a zero

## Diferenças

Teste de homogeneidade: selecionamos uma amostra de elementos de cada uma das  ***$r$  subpopulações e distribuimos os elementos de cada uma dessas amostras segundo  $x$  categorias.***

Teste de independência: distribuimos uma amostra de  **$N$  elementos de "uma" população segundo as categorias da variável A e as categorias da variável B.**

# EXERCÍCIOS

Um sociólogo afirma que a distribuição de idades dos moradores de certa cidade é diferente do que era 10 anos antes. Você seleciona aleatoriamente 400 moradores e registra a idade de cada um deles. Os resultados são registrados na tabela abaixo. Pode-se afirmar, com alfa igual a 5%, que a distribuição de idades foi alterada nesses 10 anos? E com alfa igual a 1%?

Idade	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Anterior	16%	20%	8%	14%	15%	12%	10%	5%
Pesquisa	76	84	30	60	54	40	42	14

# PRESSUPOSTOS

1º Pelo menos 5 observações por casela

2º Menos de 20% das caselas com ZERO

# ALTERNATIVAS

## → Teste Exato de Fisher

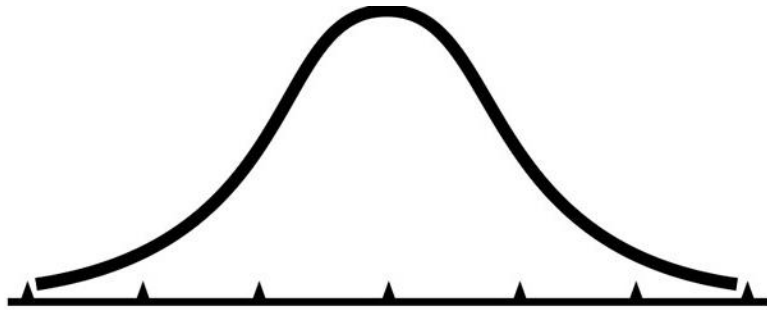
*Utiliza tabelas 2x2 e N total menor que 20 (aceita ZERO)*

## → Teste de McNemar

*Quando existe medidas “antes” e “depois”.*

## → Teste de Mantel-Haenszel

*Quando existe medidas com interferência de alguma variável associada: **Confounding***



NORMAL DISTRIBUTION



PARANORMAL DISTRIBUTION

Análise Bivariada

# TESTES NÃO-PARAMÉTRICOS

*(OUTROS)*



# Testes Paramétricos vs Testes não-paramétricos

## Testes Paramétricos:

- Baseiam-se em uma distribuição estatística subjacente dos dados
- Possuem maior poder estatístico do que os testes não-paramétricos (normalmente o p-valor de um teste paramétrico será menor do que o p-valor associado a um teste não-paramétrico associado)

## Testes Não-Paramétricos

- Não precisa seguir as condições paramétricas;
- São mais robustos (podem ser usados em uma grande gama de situações)

Testes Estatísticos			
Paramétricos		Não-Paramétricos	
Independentes	Vinculados	Independentes	Vinculados
2 amostras	2 amostras	2 amostras	2 amostras
Teste $t$ (Student)	Teste $t$ (Student)	Mann-Whitney T. da Mediana $\chi^2$ (2 x 2) Proporções Exato (Fisher)	Wilcoxon T. dos sinais Mac Nemar Binomial
Mais de duas	Mais de duas	Mais de duas	Mais de duas
Análise de variância	Análise de variância	Kruskal-Wallis Mediana (m x n) $\chi^2$ (m x n) Nemenyi	Cochran Friedman

**VER TAMBÉM:** <https://marcoarmello.wordpress.com/2012/05/17/qualteste/>

Testes Estatísticos			
Paramétricos		Não-Paramétricos	
Independentes	Vinculados	Independentes	Vinculados
2 amostras	2 amostras	2 amostras	2 amostras
Teste $t$ (Student)	Teste $t$ (Student)	Mann-Whitney	Wilcoxon
		T. da Mediana	T. dos sinais
		$\chi^2$ (2 x 2)	Mac Nemar
		Proporções	Binomial
Mais de duas	Mais de duas	Exato (Fisher)	
		Kruskal-Wallis	
		Mediana (m x n)	Cochran
		$\chi^2$ (m x n)	Friedman
Análise de variância		Nemenyi	

**VER TAMBÉM:** <https://marcoarmello.wordpress.com/2012/05/17/qualteste/>

## ***Signal test***

Pode ser aplicado em três situações:

- a. Na comparação dos resultados de amostras pareadas
- b. Na comparação dos resultados de uma amostra com mediana de uma população
- c. Na comparação de dados qualitativos

- É fácil de usar
- Praticamente não exige pressuposições
- Tem baixo poder estatístico (p-valor tem valores altos)
- Pouco utilizado

## ***Signal test***

1º Verifique quantas vezes “antes” é menor que “depois”

$\Delta (\text{depois} - \text{antes}) < 0 \rightarrow$  sinal negativo

$\Delta (\text{depois} - \text{antes}) > 0 \rightarrow$  sinal positivo

Indivíduo		1	2	3	4	5	6	7	8
Nível máximo de concentração	Antes	9	16	12	28	5	33	17	13
	Depois	14	22	18	23	11	40	15	18

-

## Signal test

1º Verifique quantas vezes “antes” é menor que “depois”

$\Delta (\text{depois} - \text{antes}) < 0 \rightarrow \text{sinal negativo}$

$\Delta (\text{depois} - \text{antes}) > 0 \rightarrow \text{sinal positivo}$

Indivíduo		1	2	3	4	5	6	7	8
Nível máximo de concentração	Antes	9	16	12	28	5	33	17	13
	Depois	14	22	18	23	11	40	15	18
SINAL		+	+	+	-	+	+	-	+

-

$$S = 6$$

$$N = 8$$

~Binomial

***Signal test***

2º Calcular as probabilidades

$$S = 6$$

$$N = 8$$

**~Binomial (N, S)**

S =	0	1	2	3	4	5	6	7	8
P (X=S)	0.0039	0.03125	0.1094	0.2188	0.2734	0.2188	0.1094	0.03125	0.0039

$$P(X = S) = \left( \frac{N!}{S!(N-S)!} \right) \cdot p^S \cdot (1-p)^{N-S}$$

## Signal test

### 3º Estabelecer Hipóteses

$H_0: \Delta(\text{mediana}) \leq 0$  (a mediana das diferenças é igual ou menor que zero)

$H_1: \Delta(\text{mediana}) > 0$  (a mediana das diferenças é maior que zero)

### 4º Estabelecer $\alpha$ ( $\alpha = 0.05$ )

### 5º Calcular o p-valor:

S =	0	1	2	3	4	5	6	7	8
P (X=S)	0.0039	0.03125	0.1094	0.2188	0.2734	0.2188	0.1094	0.03125	0.0039

p-valor = 0.0352

### 6º Concluir

Ou seja, **rejeitamos** a Hipótese Nula, pois p-valor é menor do que 5% ( $\alpha$ )



## *Signal test*



```
> pbinom(q=6,           # numero de Sucessos (sinais positivos)
+       size=8,         # N total
+       lower.tail = F, # Encontrar mais Sinais positivos do que o encontrado
+       prob=0.5)        # Para teste de sinais, consideramos que a mediana é a mesma
[1] 0.03515625
> |
```

Leva em consideração as grandezas das diferenças, assim como seus sinais.

**Procedimento:**

1. Seleccionamos uma aa. de  $n$  pares de observações
2. Calculamos a diferença de cada um dos pares ignorando os sinais dessas diferenças,  
Ordenamos seus valores absolutos do menor até o maior obtendo os postos. Uma diferença = 0 não é ordenada; ela é eliminada da análise e o tamanho da amostra é reduzido de 1 para cada par eliminado.
3. Se houver empate é atribuído um posto médio.
4. Finalmente, atribuímos a cada posto ou um sinal de mais ou um sinal de menos dependendo do sinal da diferença

**EXEMPLO**

$\bar{X}_i$	$Y_i$	$D_i$	$ D_i $	Posto	Posto sinalizado
300	350	-50	50	4	-4
410	390	20	20	1	+1
420	490	-70	70	5	-5
410	435	-25	25	2	-2
400	440	-40	40	3	-3

$H_0$ : Média das Diferenças = 0

$H_1$ : Média das Diferenças  $\neq$  0

**EXEMPLO**

$\bar{X}_i$	$Y_i$	$D_i$	$ D_i $	Posto	Posto sinalizado
300	350	-50	50	4	-4
410	390	20	20	1	+1
420	490	-70	70	5	-5
410	435	-25	25	2	-2
400	440	-40	40	3	-3

Sejam

$T^+ =$  soma dos postos positivos = 1

$T^- = -(\text{soma dos postos negativos}) = 14$

**EXEMPLO**  $T^+ = \text{soma dos postos positivos} = 1$

$T^- = -(\text{soma dos postos negativos}) = 14$

Existem  $2^5 = 32$  combinações possíveis destes 5 postos, todas elas Equiprováveis

Tabela:

1	2	3	4	5	$T^+$	1	2	3	4	5	$T^+$
+	+	+	+	+	15	+	+	-	+	-	7
-	+	+	+	+	14	-	+	-	-	+	7
+	-	+	+	+	13	-	-	+	+	-	7
+	+	-	+	+	12	+	-	-	-	+	6
-	-	+	+	+	12	+	+	+	-	-	6
+	+	+	-	+	11	-	+	-	+	-	6
-	+	-	+	+	11	+	-	-	+	-	5
+	+	+	+	-	10	-	+	+	-	-	5
-	+	+	-	+	10	-	-	-	-	+	5
+	-	-	+	+	10	+	-	+	-	-	4
-	+	+	+	-	9	-	-	-	+	-	4
-	-	-	+	+	9	+	+	-	-	-	3
+	-	+	-	+	9	-	-	+	-	-	3
+	+	-	-	+	8	-	+	-	-	-	2
+	-	+	+	-	8	+	-	-	-	-	1
-	-	+	-	+	8	-	-	-	-	-	0

**EXEMPLO**  $T^+ = \text{soma dos postos positivos} = 1$

$T^- = -(\text{soma dos postos negativos}) = 14$

Podemos testar a hipótese para  $T^+$  ou para  $T^-$

Podemos rejeitar a Hipótese Nula se:

- $T^+$  for pequeno

- $T^-$  for grande

→ Em geral utiliza-se a menor soma.

$T^+$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Freq	1	1	1	2	2	3	3	3	3	3	3	2	2	1	1	1

**EXEMPLO**  $T^+ = \text{soma dos postos positivos} = 1$

$T^- = -(\text{soma dos postos negativos}) = 14$

Calcular o p-valor.

→ Usar as probabilidades até o valor de  $T^\pm$

Por exemplo:

→ se for  $T^\pm = 2$ , conte até a  $T^\pm = 1$

→ se for  $T^\pm = 5$ , conte até a  $T^\pm = 4...$

T+	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Freq	1	1	1	2	2	3	3	3	3	3	3	2	2	1	1	1
p	1/32	1/32	1/32	2/32	2/32	3/32	3/32	3/32	3/32	3/32	3/32	2/32	2/32	1/32	1/32	1/32

**EXEMPLO**  $T^+ = \text{soma dos postos positivos} = 1$   
 $T^- = -(\text{soma dos postos negativos}) = 14$

Conclusão

$H_0$ : Média das Diferenças = 0

$H_1$ : Média das Diferenças  $\neq 0$

T+	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Freq	1	1	1	2	2	3	3	3	3	3	3	2	2	1	1	1
p	1/32	1/32	1/32	2/32	2/32	3/32	3/32	3/32	3/32	3/32	3/32	2/32	2/32	1/32	1/32	1/32



**p-valor =  $1/32 = 0,03125$   $\longrightarrow$  Rejeitamos  $H_0$**



É usado para testar se duas amostras independentes foram retiradas de populações com médias iguais.

→ Alternativa para o teste t de Student para amostras independentes quando seus pressupostos não forem contemplados (distribuição normal, N suficiente)

→ A única exigência do teste de Mann-Whitney é a de que as observações sejam medidas em escala **ordinal** ou **numérica**.

**Também chamado de Wilcoxon-Mann-Whitney ou Teste U**

# EXEMPLO

Valor	Grupo
12	Tratamento
20	Tratamento
57	CONTROLE
6	Tratamento
89	CONTROLE
71	CONTROLE
10	CONTROLE
42	Tratamento

# EXEMPLO

Valor	Grupo
6	Tratamento
10	CONTROLE
12	Tratamento
20	Tratamento
42	Tratamento
57	CONTROLE
71	CONTROLE
89	CONTROLE

# EXEMPLO

Valor	Grupo	ranking
6	Tratamento	1
10	CONTROLE	2
12	Tratamento	3
20	Tratamento	4
42	Tratamento	5
57	CONTROLE	6
71	CONTROLE	7
89	CONTROLE	8

**Grupo A:** o primeiro que aparece: Tratamento

**Grupo B:** CONTROLE

$$\mu_w = \frac{n_A(n_A + n_B + 1)}{2}$$

$$\sigma_w^2 = \frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}$$

$n_A$ : Número de elementos do Grupo A

$n_B$ : Número de elementos do Grupo B

$U$ : Soma dos valores da ordem que os elementos do grupo A ficaram

$$\mu_W = \frac{n_A(n_A + n_B + 1)}{2}$$

$$Z_W = \frac{|U - 0,5| - 0,5 \cdot n_A \cdot (n_A + n_B + 1)}{\sqrt{\frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}}}$$

$$\sigma_W^2 = \frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}$$

**$n_A$** : Número de elementos do Grupo A

**$n_B$** : Número de elementos do Grupo B

**$U$** : Soma dos valores da ordem que os elementos do grupo A ficaram

# EXEMPLO

Valor	Grupo	ranking
6	Tratamento	1
10	CONTROLE	2
12	Tratamento	3
20	Tratamento	4
42	Tratamento	5
57	CONTROLE	6
71	CONTROLE	7
89	CONTROLE	8

$$Z_w = \frac{|U - 0,5| - 0,5 \cdot n_A \cdot (n_A + n_B + 1)}{\sqrt{\frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}}}$$

$$Z_w = |13 - 0,5| - 0,5 \cdot 4 \cdot (4 + 4 + 1) / \text{RAIZ}((4 \cdot 4 \cdot (4 + 4 + 1)) / 12)$$

$$Z_w = - 1.588$$

→ Se for unicaudal, ver valor na tabela,

→ caso contrário p-valor = 2 \* p-tabela

Análise Bivariada – Estatística Aplicada

# **AVALIAÇÃO INDIVIDUAL**



Use a folha sulfite e responda ***a caneta:***

-As questões (na ordem):

- Lista teste-t

- Exercício 1
- Exercício 2
- Exercício 2 (teste t pareado)

- Lista Qui-Quadrado

- Exercício 2 (Acidentes na semana)
- Exercício 3 (Prozac)
- Exercício 4 (Três Patetas)